

From GEMBA to GPT: Artificial intelligence for the medical office assistantNathan C. Wong¹, Hanmu Yan²¹Department of Urologic Sciences, University of British Columbia, Vancouver, BC, Canada; ²St. Michael's Hospital, Toronto, ON, Canada

Cite as: Wong N, Yan H. From GEMBA to GPT: Artificial intelligence for the medical office assistant. *Can Urol Assoc J* 2026 October 19; Epub ahead of print
<http://dx.doi.org/10.5489/cuaj.9964>

Published online October 19, 2026

Corresponding author: Dr. Nathan C. Wong, Department of Urologic Sciences, University of British Columbia, Vancouver, BC, Canada; nathan.wong@medportal.ca

ABSTRACT

Administrative tasks by medical office staff are a significant but underappreciated driver of delayed patient care, staff burnout, and escalating operational costs in modern urology practice. A forced *Gemba Walk* in front-desk responsibilities exposed many previously invisible inefficiencies in my own practice, specifically the workflow with regard to faxes and patient phone calls. These tasks are high-volume, redundant, and time-consuming—making them an ideal task for artificial intelligence (AI).

Large language models (LLM) can integrate with electronic medical record (EMR) platforms to automate these repetitive tasks. For example, optical character recognition (OCR) can be used to transform incoming faxed images to operational datapoints. The models can then automatically classify the document, sort the file directly to patient charts, and clinically triage referrals and results by urgency. Furthermore, AI voice agents or AI receptionists can handle routine phone calls with full EMR integration and documentation, while complex cases can be escalated to human staff.

In an era of increasing burnout, transitioning from manual workflows to an AI-augmented paradigm allows providers to work at the top of their scope. These operational efficiencies carry broader health policy implications, including shortened referral-to-consultation intervals for urgent diagnoses, improved patient care and communication, and the generation of structured analytics to drive system-level quality improvement.

The walk: Gemba

Gemba (現場), meaning "the real place" in Japanese, is where one goes to observe and understand. In lean management philosophy, a *Gemba Walk* refers to the practice of leaders visiting the place where the work actually happens to identify opportunities for systematic improvement.¹ In medicine, the Gemba of clinical administration is seldomly visited by the physician. Often, we are too busy reviewing referrals, treating patients, and dictating notes, rarely witnessing the upstream processes that keep a medical practice running.

For me, this all changed on a Monday morning in March when my sole medical office assistant (MOA) called in sick. I was alone. With no last-minute emergency coverage available and a full clinic ahead, I had one of two options: 1) cancel my clinic and reschedule all of the patients to a later date or 2) act as both the MOA and urologist. Regrettably, or now fortunately, I chose the latter. What followed was one of the most instructive weeks of my career, not in urology, but in the hidden operational mechanics of a busy medical practice.

The fax machine was my first nemesis. Since its introduction in the 1980's, the fax machine has remained the predominant method of inter-provider communication. As I stepped through the clinic door, the faxes were already piled up. At 8 am, 31 faxes were in queue: 7 referrals, 12 laboratory results, 5 imaging reports, 3 pharmacist prescription requests, and 4 consultation reports. Each fax required me to manually open each of the PDF files and analyze the report individually. Who sent this fax? Is this patient in my system? Is this a result or a new referral? Is this urgent? Where does it go? The cognitive structure of this task is deceptively simple but unrelentingly repetitious requiring countless clicks of the mouse, being performed 31 times before the clinic even opened.

At 9 am, the telephone became my next adversary. From open to close, 64 calls were received. Approximately 80% followed one of five scripted scenarios: 1) appointment booking or cancellation, 2) office information, 3) prescription refill requests, 4) requisition update, and 5) referral status. Only 8 required genuine clinical judgment, including post-op care and medication side effects. By Tuesday, overwhelmed and fatigued, I had switched my phone line to go directly to voicemail and had made 2 routing errors: a lab result filed to the wrong chart and a prescription refill left unsent until the patient called back again. These were not errors of negligence, but errors of volume, monotony, and cognitive overload. By Wednesday, I understood with visceral clarity what my MOA does for 40 hours a week. There must be another way.

The talk: GPT

The tasks I performed that week share 3 characteristics ideal for AI automation: 1) they are high-volume, 2) they follow recognizable patterns, and 3) they carry meaningful consequences when performed incorrectly. LLMs such as GPT-4 are able to read unstructured text, extract key data points (name, dates, diagnoses), classify documents, and generate structured outputs compatible with EMR data fields.²

Incoming faxes received as PDF image files would be processed through optical character recognition (OCR) to convert them to machine-readable text. This technology is mature, widely deployed, and achieves greater than 94% accuracy on typed medical documents.³ The OCR output is then passed to an LLM-based classification pipeline with clear functions.

First, the document requires classification. The AI model assigns each document to a category (referral, lab result, imaging, OR booking, admin, correspondence, consult, other) with a confidence score. Documents below a confidence threshold are flagged for human review rather than processed automatically, preserving a human-in-the-loop safety net.

Data entity extraction comes next. The AI model identifies and extracts structured data fields including patient name, date of birth, personal health number, referring provider name and fax number, clinical indication, urgency signals, and required action. These fields are populated into EMR-compatible structured data formats, bypassing manual transcription and the associated error risk entirely.

Subsequently, faxes are matched to patient charts. Using the extracted identifiers, the AI queries the EMR to determine whether the patient exists in the current database. Confirmed matches route the document automatically to the appropriate chart while unmatched patients trigger the automatic creation of a new patient chart, pre-populated with the extracted data points, requiring only verification by the MOA rather than full manual entry.

Finally, the files are triaged and routed. Based on clinical content, the model assigns an urgency tier of the referral or result consistent with referral prioritization guidelines.⁴ For example, Tier 1 or “Urgent” conditions (e.g., suspected malignancy, obstructive uropathy, gross hematuria) are immediately pushed to the top of the list by notifying the urologist via EMR alert. Subsequent lower-tier items are queued for next-available booking guided by set parameters of the provider.

For telephone management, AI receptionists operate through a different but complementary mechanism. Intelligent voice agents, distinct from interactive voice response systems in their ability to engage in natural-language dialogue, can handle the 56 routine call types in my Gemba walk. Integration with the EMR allows these AI agents to verify patient identity, review office details (location, fax number, hours, parking), pull chart context, process appointment changes, route prescription refill requests to the prescribing provider's queue with a structured summary, and answer protocolized post-operative questions from a pre-approved knowledge base. Calls outside these parameters are escalated in real-time to the “human” MOA, with a transcribed call summary by the AI to assist the staff with the patients’ concerns. Furthermore, all AI-generated actions are logged with timestamps and decision rationale, creating an auditable trail that most manual workflows lack.

These proposed AI workflows have been successively implemented in clinical practice, albeit in the early phases with ongoing modifications and updates. Failures do occur, thus necessitating human oversight. Errors that have already been noted include misclassifying notes, misfiling lab or imaging results in the incorrect patient charts, and creating new charts with

incorrect demographic information (such as switching first and last names). In the largest clinician-annotated evaluation of LLM clinical note generation to date — nearly 13,000 sentences across 450 notes, reviewed by 50 physicians — an LLM pipeline produced a 1.47% hallucination rate and a 3.45% omission rate.⁵ A separate simulation study of two commercial ambient scribes found an average of 2.9 errors per note, most often omissions.⁶ With experience and ongoing learning of the LLM, these errors are expected to decrease over time, though they are unlikely to ever reach zero.

This is why verification cannot be an afterthought. When a provider cannot quickly confirm an AI-generated claim, the work of verification does not disappear — it moves, and AI output that has to be checked may increase cognitive burden rather than decrease it.⁷ The concern is compounded by automation bias: under time pressure, reviewers tend to accept plausible AI output without close scrutiny.⁸ These failures share a common root; AI often acts without a visible link back to the source data. This points to a design principle beyond human review — each extracted datapoint should ideally be directly anchored to the region of the source data it was derived from, so that human verification can be rapid. Traceability makes an automated action checkable instead of opaque, and lets a confidence threshold fail safely rather than silently.

With any discussion of AI in medicine, confidentiality and privacy must be addressed directly. The AI software is PIPEDA and, in Ontario, PHIPA, compliant with Canadian-hosted data infrastructure and a prior privacy impact assessment. Under PHIPA, the physician remains the health information custodian, and any software processing that information acts as the custodian's agent, on the custodian's authority and under a written agreement.⁹ Notably, Canadian data residency is a risk-mitigation choice rather than a statutory requirement. Neither PIPEDA nor PHIPA requires domestic data storage; instead, both require reasonable safeguards, accountability, and transparency about information practices, and both permit processing by a third party where a comparable level of protection is ensured by contract.⁹⁻¹¹ Because most commercial health-AI infrastructure relies on US-parent cloud providers, the US Clarifying Lawful Overseas Use of Data (CLOUD) Act can compel those providers to produce data regardless of server location. To manage this structural risk, organizations must use Canadian region-pinning, implement data-processing agreements, practice data minimization, and explicitly disclose residual risks to custodians.¹¹

Patient acceptance of AI-mediated interactions is also not uniform. Older adults or those with complex psychosocial concerns may find AI receptionists impersonal or difficult to navigate, and an accessible escalation pathway to a human staff member must always be preserved.

Beyond individual practice, AI-driven automation carries health policy implications across patient access to care, provider burnout, and data quality improvement. Automation of routine tasks preserves humans for more complex tasks, allowing the MOA role to evolve to involve more direct patient care. Autonomous triage shortens the referral-to-consultation interval

for time-sensitive diagnoses and, compared to written or manually transcribed entries, AI-generated notes are inherently more structured, consistent, and analyzable. These data points can populate registries, inform waitlist analytics, and support system-level quality improvement.

The view from the front desk is irreplaceable. The administrative scaffolding of modern urologic care is labor-intensive, repetition-heavy and error-prone under volume, all largely invisible to the practicing physicians. The technology exists. The clinical need is acute. The workforce pressures are real. What is required is the collective will to bring AI-driven workflow automation from proof of concept to standard of practice.

DRAFT

REFERENCES

1. Mai M. Gemba Kaizen: A commonsense approach to a continuous improvement strategy. 2nd ed. McGraw-Hill; 2012.
2. Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. *Nature* 2023;620:172-80. doi:10.1038/s41586-023-06291-2
3. Ferrando J, Domínguez JL, Torres J, et al. Improving accuracy and speeding up document image classification through parallel systems. *Computational Science - ICCS* 2020. 2020;12138:387-400. doi:10.1007/978-3-030-50417-5_29
4. Taranaki District Health Board. Urology planned care services referral guidelines. Taranaki District Health Board. New Zealand; 2022. Available from: <https://coilink.org/20.500.12592/4of06xf>
5. Asgari E, Montaña-Brown N, Dubois M, et al. A framework to assess clinical safety and hallucination rates of LLMs for medical text summarisation. *NPJ Digit Med* 2025;8:274. doi:10.1038/s41746-025-01670-7
6. Biro JM, Handley JL, Mickler J, et al. The value of simulation testing for the evaluation of ambient digital scribes: A case report. *J Am Med Inform Assoc* 2025;32:928-31. doi:10.1093/jamia/ocaf052
7. Bedard J, Kropp M, Hsu M, et al. When using AI leads to “brain fry.” *Harvard Business Review*. March 5, 2026.
8. Qazi IA, Ali A, Khawaja AU, et al. Automation bias in large language model-assisted diagnostic reasoning among physicians trained in AI literacy: A randomized clinical trial. *NEJM AI* 2026;3. doi:10.1056/AIoa2501001
9. Personal Health Information Protection Act, 2004, S.O. 2004, c. 3, Sched. A, ss. 6, 12, 16, 17, 50.
10. Personal Information Protection and Electronic Documents Act, S.C. 2000, c. 5, Schedule 1, Principle 4.1.3.
11. Office of the Privacy Commissioner of Canada. Guidelines for processing personal data across borders. 2009.

DISCLOSURES. Dr. Yan is the founder of Translational AI Health Solutions Inc., which develops source-anchored clinical documentation software of the kind discussed in this article. He received no funding for this work and had no role in the software deployment described by the lead author.